

14.04.00
Substitution sheetTitle

A Method and an Arrangement for Objective Assessment
of Video Quality.

5

Field of the Invention

The present invention relates to video quality and, in particular, to an objective assessment of the quality of coded and transmitted video signals.

10

Background of the Invention

With the development of digital coding technology savings in transmission and/or storage capacity of video signals have been achieved and a large number of new multi-media video services have become available.

15

Savings in transmission and/or storage capacity by digital compression technology generally depend upon the amount of information present in the original video signal, as well as how much quality the user is willing to sacrifice. Impairments may result from the coding technology used and limited transmission channel capacity.

20

Video quality assessment can be split into subjective assessment by human observers providing their subjective opinion on the video quality, and objective assessment which is accomplished by use of electrical measurements.

25

It is the general opinion that assessment of video quality is best established by human observers which is, however, a complex, costly and time consuming approach. Accordingly, there is a need to develop objective visual quality measures, based on human perception, that can be used to predict the subjective quality of modern video services and applications.

Studies in the framework of the American National Standards Institute (ANSI) and the International Telecommunication Union (ITU) have led to a plurality of algorithms for objective video quality assessment.

30

As will be appreciated by those skilled in the art, calculation of quality indicators of video signals on a pixel bases, for example, requires a large amount of processing. As disclosed in a conference publication by S.D. Voran "The development of objective video quality measures that emulate

human perception", Globocom'91 conf. publ. vol. 3, pp. 1776-1781, 1991, an important class of disturbing distortions in a video signal are those that destroy, soften, blur, displace, or create edges or signal transitions in the video image.

In a further conference publication by S.D. Voran and S. Wolf "An objective technique for assessing video impairments", IEEE Pacific RIM Conference on Communications, Computers and Signal Processing, Proceedings Volume 1 of 2, pp 161-165, 1993, an objective technique is described, which is based on digital image processing operations performed on digitized original and impaired video sequences. The technique implies a features extraction process in which so called impairment measurements of perceptual video attributes in both the spatial and temporal domains are determined. The spatial impairment measurement is based on a Sobel filtering operation or, alternatively, a "pseudo-Sobel" operation, in order to enhance the edge content in the video image, and consequently in the spatial impairment measurement. The spatial impairment measurement is based on normalised energy differences of the Sobel-filtered video frames using standard deviation calculations conducted over visible portions of the pixel arrays of the original and impaired video signals. The impairment measurements thus extracted from the original and impaired video sequences are then used to compute a quality indicator that quantifies the perceptual impact of the impairments present in the impaired video sequence. The patent publication US-A-5,446,492 discloses a similar technique in which the feature-extraction processes on the original and impaired video sequences are carried out at distantly apart source and destination locations. The features extracted from the original video sequence are such that they can be easily and quickly communicated between the source and destination locations via a separate low-bandwidth transmission path, i.e. the bandwidth of the source features is much less than the bandwidth of the original video sequence. To this end the feature-extraction process additionally includes a statistical subprocess which subjects the output of the Sobel filtering operation to a statistical processing, i.e. the computation of the standard deviation of the pixels contained within a so called region of interest for which the video quality is to be measured.

A drawback of these known techniques is the fact that the feature-extraction process is based on standard deviation calculations. One thing and another means that image distortions having contrary effects in the Sobel frames, e.g. blurring vs additional noise or false edges, can not always be detected. A further drawback is that the known techniques use a relative distance measure for the quality of perception, which consequently is sensitive for relative effects of very small size and as such of small visibility.

Summary of the Invention

The present invention aims to provide objective quality measures that can be used to assess the subjective quality of video signals, dealing with the higher levels of cognitive processing which dominate the perception of video quality.

It is a further object of the present invention to provide such measures applicable for standardisation.

It is a still further object of the present invention to provide a method, an arrangement and equipment for objective quality assessment of degraded video signals for measuring the quality of video coding equipment and algorithms, video transmissions and other multi-media video services, , and which among other things do not have the above mentioned drawbacks.

These and other objects and features are achieved by the present invention in a method of obtaining quality indicators for an objective assessment of a degraded or output video signal with respect to a reference or input video signal by quantifying the strength of edges or signal transitions in both the input and the output video signals using edge or signal transition detection, which method comprises a first main step of generating image features of the input and output video signals, and a second main step of determining quality indicators from the generated image features, and for the definition of which method the prior art of document US-A-5,446,492 has been used. The process of quantifying the strength of the edges will hereinafter be referenced by the term edginess.

The method according to the invention includes in the first main step the steps of:

- a) detecting edges in the input and the output video signals; and
- b) calculating the edginess of the input and the output video signals, providing input and output edge signals; and

in the second main step the steps of

- c) establishing introduced edges in the output edge signal by comparing the input and output edge signals of corresponding parts of the input and output video signals, introduced edges being edges which are present in the output edge signal and are absent at corresponding positions in the input edge signal;

- d) establishing omitted edges in the output edge signal by comparing the input and output edge signals of corresponding parts of the input and output video signals, omitted edges being edges which are present in the input edge signal and are absent at corresponding positions in the output

edge signal;

e) obtaining normalised values of the introduced edges relative to the output edge signal adjusted by a first normalisation factor;

f) obtaining normalised values of the omitted edges relative to the input edge signal adjusted by a second normalisation factor;

g) calculating a first quality indicator by averaging the values obtained in step e); and

h) calculating a second quality indicator by averaging the values obtained in step f).

The method according to the invention is based on human visual perception, characterised in that spatial distortions like the introduction and omission of edges or signal transitions have a great impact on the subjective quality of the video signal. Further, it has been found that the introduction of an edge is more disturbing than the omission of an edge.

This has been taken into account, in the method according to the invention, by obtaining normalised values of the introduced edges and the omitted edges. The introduced edges are normalised with respect to the output edge signal adjusted by a first weighing or normalisation factor and the omitted edges are normalised with respect to the input edge signal adjusted by a second weighing or normalisation factor. Obtaining normalised values according to the present invention is more in line with human perception, which is always relative.

The quality indicators for both the introduced and the omitted edges are subsequently established by calculating mean values of the thus normalised introduced and omitted edges or signal transitions in the output video signal.

For a number of different types of video signals, classified by the amount of motion in the pictures, the quality indicators obtained with the invention are close to the quality indicators obtained from subjective measurements by human observers.

In a preferred embodiment of the method according to the invention, the proportions of introduced and omitted edges are established from respective polarities of a bipolar distortion signal formed from difference building of aligned, corresponding unipolar input and output edge signals of corresponding parts of the input and output video signals.

The first and second normalisation factors may be fixed or, preferably, set in accordance with the characteristics of the video signals, such as the luminance and chrominance values thereof.

For high luminance values, edge deteriorations are less visible which, in a further embodiment of the invention, is taken into account in that the first normalisation factor comprises a variable part obtained from the maximum characteristic values of the video signals, such as the luminance signal.

Calculation of the edginess can be established in a variety of manners. However, the most straightforward mathematical formulation is to calculate the norm of the gradient of the video signals. An example hereof is Sobel filtering which has proven to provide reliable results. Depending on how derivatives of the video signals are approximated, many variations in the calculation of the edginess are feasible. All these types hereinafter will be referred to as Sobel filtering.

In a preferred embodiment of the invention, in particular wherein the introduced and omitted edges are obtained from a distortion signal formed from aligned input and output edge signals, improved or smeared Sobel filtering provides excellent results. With smeared Sobel filtering, a smearing operator having a width of, for example, 3 pixels is used. By this smearing operation, the effect of misalignment in the formation of the distortion signal is compensated for.

Alignment of the input and output edge signals is required because video sequences processed by a codec or transmitted over a transmission channel, for example, show delays with respect to the original sequence and which vary from picture to picture. If the video sequence contains relative little motion, there is only a little influence on the objective video quality measure. However, with large movements the omission of delay compensation leads to a large mismatch in scene content between original and distorted sequences. This inadvertently increases the computed distortions. To solve the time varying delay problem, known alignment algorithms can be used such as disclosed by ITU-T Contribution COM-12-29, "Draft new recommendation on multi-media communication delay, synchronisation, and frame rate measurement", Dec. 1997.

In practice, in accordance with the invention, the quality indicators are obtained from the luminance and chrominance representations of a colour video signal.

Heuristic optimisation has led to quality indicators obtained from smeared Sobel edge detection wherein for the luminance signals the constant part of the first normalisation factor is in a range between 15 and 30, preferably 20; the constant part of the second normalisation factor is in a range between 5 and 15, preferably 10; and the variable part of the first normalisation factor is in a range between 0.3 and 1, preferably 0.6 times the maximum edge values of the luminance signal of the input and output video signals. For the chrominance signals, the constant part of the first and second weighing factors is in a range between 5 and 15, preferably 10.

From the thus obtained first and second quality indicators of each the luminance and chrominance signals, weighted quality indicators are obtained. For example, using multiple linear regression techniques. For a Mean Opinion Score (MOS) calculated from the weighted quality indicators obtained from the above smeared Sobel filtering and preferred weighing factors, correlation of the calculated MOS and observed MOS from subjective measurements reaches a value of above 0.9 which is required for making reliable predictions.

The best results are obtained from training the method on subjective reference quality data such that the normalisation factors and/or weighing of the quality indicators are optimised.

The invention further provides an arrangement for obtaining quality indicators for an objective assessment of a degraded or output video signal with respect to a reference or input video signal by quantifying the strength of edges or signal transitions in both the input and the output video signals using edge or signal transition detection, which arrangement comprises means for generating image features of the input and output video signals and means for determining quality indicators from the generated image features, for the definition of the arrangement the document US-A-5,446,492 has been used. The arrangement according to the invention includes in the means for generating image features:

a) means for detecting edges in the input and the output video signals; and

b) means for calculating the edginess of the input and the output video signals, providing input and output edge signals;
and in the means for determining quality indicators:

c) means for establishing introduced edges in the output edge signal by comparing the input and output edge signals of corresponding parts of the input and output video signals, introduced edges being edges which are present in the output edge signal and are absent at corresponding positions in the input edge signal;

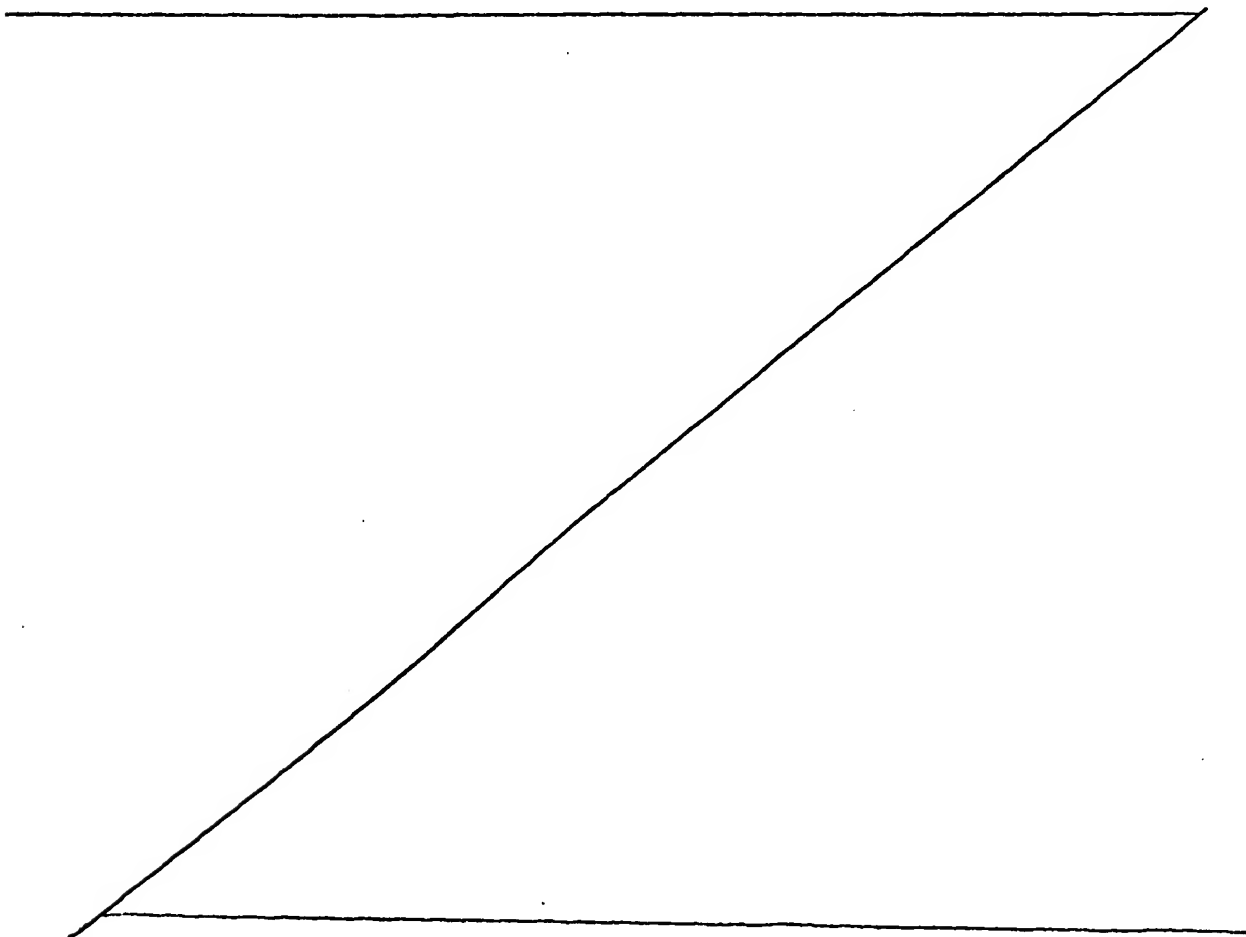
d) means for establishing omitted edges in the output edge signal by comparing the input and output edge signals of corresponding parts of the input and output video signals, omitted edges being edges which are present in the input edge signal and are absent at corresponding positions in the output edge signal;

e) means for obtaining normalised values of the introduced edges relative to the output edge signal adjusted by a first normalisation factor;

f) means for obtaining normalised values of the omitted edges relative to the input edge signal adjusted by a second normalisation factor;

6a : : (substitution sheet)

The above and other features and advantages of the present invention will be readily apparent to one of ordinary skill in the art from the following written description when read in conjunction with the drawings in which like reference numerals refer to like elements.



Brief Description of the Drawings

Fig. 1 shows an approach towards objective measurement of the quality of a video system wherein a perceptual/cognitive model simulates a human subject.

Fig. 2 shows, in a schematic representation, basic objective measurement following the approach shown in fig. 1.

Fig. 3a shows a video type picture and fig. 3b shows edges or signal transitions in the picture of fig. 3a.

Fig. 4 shows, in a schematic and illustrative manner, an exemplary embodiment of establishing introduced and omitted edges in an output video signal.

Fig. 5 shows a flow chart type diagram of the main embodiment of the method according to the invention.

Fig. 6 shows a block diagram of an arrangement according to the present invention.

Detailed Description of the Embodiments

Two main categories of video quality assessment can be distinguished. That is subjective assessment and objective assessment. Objective assessment of video quality, such as for television broadcast, in the past has been accomplished through resolution charts, colour bars, frequency response measurements, and by measuring the signal to noise ratio of the video signal. With the introduction of advanced digital video coding and decoding techniques, classical objective quality measurements like signal to noise ratio and total harmonic distortion have been proven less useful. These classical objective visual quality measurements do not take the user into account who decides by its own subjective assessment whether a video signal meets an expected quality level. The availability of objective test methods that show a high correlation with the subjective quality is therefore important.

Fig. 1 shows a so-called black box approach towards objective measurement of the quality of a video system 1, such as a video codec (coder/decoder), a transmission device or transmission path etc.

A video input or reference signal Y_{in} provided at an input terminal 2 is processed by the video system 1 into an output video signal Y_{out} at an output terminal 3.

5 A human subject 4 receiving the output video signal Y_{out} through his eyes, will not perceive any differences between the input and output video signals for an ideal video system 1. However, in practice, the video system 1 will degrade the output compared to the input, leading to a quality impression by the human subject. The level of the quality impression is determined both by the perception of the input and output
10 signals by the human subject, i.e. his visual organs, as well as his cognition, i.e. the manner in which the perceived signals are interpreted by his brains.

Accordingly, objective quality assessment of video signals correlating with subjective quality impressions, has to be based
15 on both a perceptual and cognitive model 5.

Fig. 2 shows the basic principles of objective measurement, wherein a perceptual model 6 transforms the input and output video signals into an internal representation of the input and output video signals. The difference 7 in the internal representations is mapped at
20 a quality level using a cognitive model 8. Perceptual models 6 are generally based on edge or signal transition, using spatial filters originating from the human visual system.

It is the aim of the present invention to assess the quality perception of a human subject in an as much as possible accurate
25 manner from objective measurements.

Fig. 3a shows a video type picture, whereas fig. 3b shows edges or signal transitions detected in the picture. For illustration purposes, an arbitrary corresponding part in both figures is indicated by an ellips 9.

30 The invention is primarily based on edge or signal transition detection of a video signal as illustrated, by way of example, in fig. 4.

Signal parts 10 and 11 of an input video signal Y_{in} and an output video signal Y_{out} , respectively, are time and space aligned in
35 order to rule out inherent time delays or spatial shifts in the video system 1 (fig. 1). The signal part 10 comprises an edge or signal transition 12, as shown, while signal part 11 comprises an edge or signal

transition 13, shifted in position with respect to edge 12 and of a lower amplitude. In the present example, edge 13 is assumed to be a new or introduced edge in the output signal Y_{out} .

In the ideal undistorted case, both signal parts 11 and 12 have edges at the same position and of the same amplitude, such that they cancel each other in a distortion signal formed by difference building.

The signal parts 10 and 11 are subjected to an edge operator, indicated by an arrow 14, for quantifying the strength of the edges 12 and 13, providing edge signals X_{in} and X_{out} , respectively, referenced by numerals 15 and 16.

As shown, the edge signals are of a unipolar nature, i.e. for both the leading and trailing portions of the edges 12 and 13 correspondingly positioned positive pulses 17 and 18 are provided.

In a next step, indicated by an arrow 19, a distortion signal 20 is formed by subtraction of the edge signal 15 from the edge signal 16, i.e. $(X_{out} - X_{in})$. This distortion signal 20 is of a bipolar nature, as shown. From the distortion or difference signal 20 introduced and omitted edges are established as indicated by arrows 21 respectively 22.

The positive part of the bipolar distortion signal 20, i.e. noted $(X_{out} - X_{in})^+$, provides introduced edges in the output signal 11. That is, the edge 13 which is not present in the input signal. The negative portion of the distortion signal 20, i.e. noted $(X_{out} - X_{in})^-$, provides the omitted edges in the output signal 11, that is the edge 12 of the input signal 10 which is not present in the output signal 11.

Those skilled in the art will appreciate that edge detection for the purpose of the present invention can be established by a differentiation or derivative operation, generating a plurality of edge magnitude values based on the image values, i.e. referenced as edginess.

In a preferred embodiment of the invention, the edge detector can be a Sobel operator that generates the magnitude values from the derivative of the image values, i.e. the rate of change of intensity over a plurality of image pixels.

Sobel filtering is an operation which is well known in the art of image processing and can be found in many text books on this

subject, such as "Digital Image Processing", by R.C. Gonzalez and P. Winz 2nd Ed.; Addison-Wesley Publishing Co., Reading, Massachusetts, 1987.

In the classical Sobel filtering, the rate of change of the pixels of a video image along the x-axis, i.e. the horizontal edges, is determined by convolving the video image with the matrix:

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

The rate of change of the pixels of the video image along the y-axis, i.e. the vertical edges, is determined by convolving the image with the matrix:

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

The square root of the sum of both edge detectors provides the edge magnitude in a pixel or point of the video image.

The invention makes use of the insight that human perception is a relative operation, such that a weighing or normalization of the introduced and omitted edges is established.

Accordingly, the introduced edges are normalized relative to the output edge signal adjusted by a first normalization factor and the omitted edges are normalised relative to the input edge signal adjusted by a second normalization factor. This leads to the following equations:

$$Q1 = \text{mean} \frac{(X_{\text{out}} - X_{\text{in}})^+}{X_{\text{out}} + W1} \quad (1)$$

$$Q2 = \text{mean} \frac{(X_{\text{out}} - X_{\text{in}})^-}{X_{\text{in}} + W2} \quad (2)$$

wherein:

Q1 = first quality indicator for introduced edges.

Q2 = second quality indicator for omitted edges.

W1 = first normalization factor.

W2 = second normalization factor.

5 The quality factors are separately calculated for the luminance and chrominance parts of colour video signals. The normalization factors W1 and W2 are set in accordance with the characteristics of the video signals and may comprise a constant part corresponding to luminance and chrominance values of the video signals.

10 In a preferred embodiment of the invention, the first normalization factor comprises a variable part obtained from the maximum edge values of the input and output video signals. Preferably, the maximum edge values of the luminance signal of the input and output video signals. This, because edge deteriorations for high luminance values are less visible. Applying Sobel operation as the edge operator in accordance with
15 the invention, equations (1) and (2) can be written as:

$$20 \quad Q1 = \text{mean} \frac{\{\text{Sobel}(Y_{\text{out}}) - \text{Sobel}(Y_{\text{in}})\}^+}{\text{Sobel}(Y_{\text{out}}) + W1} \quad (3)$$

$$25 \quad Q2 = \text{mean} \frac{\{\text{Sobel}(Y_{\text{out}}) - \text{Sobel}(Y_{\text{in}})\}^-}{\text{Sobel}(Y_{\text{in}}) + W2} \quad (4)$$

wherein:

Sobel(Y) is Sobel filtering of the video signal Y.

30 As discussed above, for the purpose of the present invention, the input and output video signals have to be time and space aligned.

35 In order to correct for misalignments and to adapt the processing with respect to the spatial resolution of the video signal, the Sobel filter has been enhanced by a so-called Smearing operator, having a width of a few pixels. Use of this smearing operator has the effect of extending the Sobel filter operation over the image pixels.

A three pixel width smearing of the filtered signal is defined as:

$$\text{Smeared Sobel } (Y) = \text{MAX} \left\{ i = -1, 0, 1 \quad j = -1, 0, 1 \quad \text{Sobel }_{i,j} (Y) \right\} \quad (5)$$

wherein:

i, j = pixels in x and y directions over which the Sobel filtering is extended.

Again, for reliable results, first and second quality indicators for introduced and omitted edges have to be separately calculated for the luminance and chrominance parts of a colour video signal.

From heuristic optimization, for the luminance signals, reliable results are provided by a first normalization factor $W1$ in a range between 15 and 30, preferably 20 and comprising a variable part in a range between 0.3 and 1, preferably 0.6 times the maximum edge values of the luminance signal, and a second normalization factor in a range between 5 and 15, preferably 10. For the chrominance signals, the first and second normalization factors are to be chosen in a range between 5 and 15, preferably 10.

Accordingly, for the luminance signals, in a preferred embodiment of the invention excellent quality indicators are obtained from:

$$Q1(L) = \text{mean} \frac{\{\text{Smeared Sobel } (Y_{out}^L) - \text{Smeared Sobel } (Y_{in}^L)\}^+}{\text{Smeared Sobel } (Y_{out}^L) + 10 + 0.6 \text{ MAX } (X_{in}^L, X_{out}^L)} \quad (6)$$

$$Q2(L) = \text{mean} \frac{\{\text{Smeared Sobel } (Y_{out}^L) - \text{Smeared Sobel } (Y_{in}^L)\}^-}{\text{Smeared Sobel } (Y_{in}^L) + 10} \quad (7)$$

wherein:

$Q(L)$ = quality indicator for luminance signals

Y^L = luminance signal

X^L = edge luminance signal

The rational for the factor $0.6 \text{ MAX } (X_{in}^L, X_{out}^L)$ lays in the Weber law, which states that subjects are less sensitive to absolute contrast for large luminance values.

For both chrominances signals C_R and C_B of the video signals, quality indicators are obtained from:

$$Q1(C_R) = \frac{\{\text{Smeared Sobel } (Y_{out}^{C_R}) - \text{Smeared Sobel } (Y_{in}^{C_R})\}^+}{\text{Smeared Sobel } (Y_{out}^{C_R}) + 10} \quad (8)$$

$$Q2(C_R) = \frac{\{\text{Smeared Sobel } (Y_{out}^{C_R}) - \text{Smeared Sobel } (Y_{in}^{C_R})\}^-}{\text{Smeared Sobel } (Y_{in}^{C_R}) + 10} \quad (9)$$

$$Q1(C_B) = \frac{\{\text{Smeared Sobel } (Y_{out}^{C_B}) - \text{Smeared Sobel } (Y_{in}^{C_B})\}^+}{\text{Smeared Sobel } (Y_{out}^{C_B}) + 10} \quad (10)$$

$$Q2(C_B) = \frac{\{\text{Smeared Sobel } (Y_{out}^{C_B}) - \text{Smeared Sobel } (Y_{in}^{C_B})\}^-}{\text{Smeared Sobel } (Y_{in}^{C_B}) + 10} \quad (11)$$

wherein:

$Q(C_R)$ = quality indicator for chrominance C_R signal.

$Q(C_B)$ = quality indicator for chrominance C_B signal.

Y^{C_R} = chrominance C_R signal.

Y^{C_B} = chrominance C_B signal.

Using multiple linear regression techniques, a Mean Opinion Score (MOS) can be calculated from the six quality indicators obtained and three quality indicators derived from the LP-distance between the reference or input video signal and the degraded or output video signal. With the present invention, correlation of the calculated MOS and the observed MOS from subjective measurements of human test persons reaches a value of above 0.9 which is required for standardisation purposes within ANSI-ITU.

The normalization factors and weighing of the quality indicators can be further optimized from running the method according to

the invention for subjective quality data for a number of training video sequences, for example.

Figure 5 shows, in a flow chart type diagram, the main steps of the method according to the present invention.

5 Detection of edges and calculation of the edginess of the input video signal are schematically indicated by blocks 25 and 26, while blocks 27 and 28 disclose same for the output video signal.

10 From the edge signals obtained in the blocks 26 and 28, introduced and omitted edges in the output signal are established, referenced by blocks 29 and 33, respectively.

15 In accordance with human perception, normalised values are obtained, by introducing a first and second normalization factor as indicated by blocks 30 and 34, respectively. The introduced edges are normalized with respect to the output edge signal, whereas the omitted edges in the output signal are normalized with respect to the input edge signal.

20 Averaging of the values obtained, blocks 31 and 35, leads to the first and second quality indicators according to the present invention, referenced by blocks 32 and 36, respectively.

Figure 6 shows a schematic block diagram of an arrangement for obtaining quality indicators for an objective assessment of video signal quality in accordance with the method of the invention.

25 At input terminal 40 an output video signal Y_{out} is provided, time and space aligned with an input video signal Y_{in} , provided at input terminal 41. Edge detection and calculation means 42 and 43 are operative to detect edges and to quantify the strength of the edges, i.e. the edginess of the respective video signals, providing output and input edge signals, respectively. In a preferred embodiment of the invention, the means 42 and 43 are arranged for Sobel filtering of the video signals, in particular smeared Sobel filtering.

30 The edge signals provided are fed to means 44 for establishing a distortion signal or difference signal, such as the distortion signal 20 illustrated in figure 4.

35 Detection means 45 operate on the distortion signal for detecting introduced edges in the output edge signal and detection means 46 operate on the distortion signal to detect omitted edges in the output edge signal.